Scientific Committee on Occupational Exposure Limits (SCOEL). 2012. Recommendation from the Scientific Committee on Occupational Exposure Limits for Ethylene Oxide, SCOEL/SUM/160.

Shore, R, Gardner, M, Pannett, B. 1993. Ethylene oxide an assessment of the epidemiological evidence on carcinogenicity. British Journal of Industrial Medicine. 50:971-997.

Sielken Jr, R, Valdes-Flores, C. 2009a. Calculating excess risk with age-dependent adjustment factors and cumulative doses ethylene oxide case study. Regulatory Toxicology and Pharmacology. 55:76-81.

Sielken Jr, R, Valdes-Flores, C. 2009b. Life-table calculations of excess risk for incidence versus mortality. Regulatory Toxicology and Pharmacology. 55:82-89.

Stayner, L, Steenland, K, Greife, A, Hornung, R, Hayes, R, Nowlin, S, Morawetz, J, Ringenburg, V, Elliot, L, Halperin, W. 1993. Exposure response analysis of cancer mortality in a cohort of workers exposed to ethylene oxide. American Journal of Epidemiology. 138:787-798.

Steenland, K, Stayner, L, Greife, A, Halperin, W, Hayes, R, Hornung, R, Nowlin, S. 1991 Mortality among workers exposed to ethylene oxide. The New England Journal of Medicine. 324(20):1402-1407.

Steenland, K, Whelan, E, Deddens, J, Stayner, L, Ward, E. 2003. Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States). Cancer Causes and Control. 14:531-539.

Steenland, K, Stayner, L, Deddens, J. 2004. Mortality analyses in a cohort of 18 235 ethylene oxide exposed workers: follow up extended from 1987 to 1998. Occup Environ Med. 61:2-7.

Steenland, K, Seals, R, Klein, M, Jinot, J, Kahn, H. 2011. Risk Estimation with Epidemiologic Data When Response Attenuates at High-Exposure Levels. Environmental Health Perspectives. 119(6):831-837.

Swaen, G, Slangen, J, Ott, M, Kusters, E, Van Den Langenbergh, G, Arends, J, Zober, A. 1996. Investigation of a cluster of ten cases of Hodgkin's disease in an occupational setting. Int Arch Occup Environ Health. 68:224-228.

Swaen, G, Burns, C, Teta, J, Bodnar, K, Keenan, D, Bodnar, C. 2009. Mortality study update of ethylene oxide workers in chemical manufacturing: a 15 year update. JOEM. 51(6):714-723.

Swenberg, JA, Fryar-Tita, E, Jeong, YC, Boysen, G, Starr, T, Walker, VE, & Albertini, RJ. 2008. Biomarkers in toxicology and risk assessment: informing critical dose-response relationships. Chem Res Toxicol 21(1):253-265.

Teta, M, Benson, L, Vitale, J. 1993. Mortality study of ethylene oxide workers in chemical manufacturing a 10 year update. British Journal of Industrial Medicine. 50:704-709.

Teta, J, Sielken Jr, R, Valdez-Flores, C. 1999. Ethylene oxide cancer risk assessment based on epidemiological data application of revised regulatory guidelines. Risk Analysis. 19(6):1135-1155.

TCEQ. 2015. Guidelines to develop toxicity factors. RG-442: Texas Commission on Environmental Quality (TCEQ).

United States Environmental Protection Agency (USEPA). 2005a. Guidelines for Carcinogen Risk Assessment. (EPA/630/P-03/001B). Washington, D.C.

United States Environmental Protection Agency (USEPA). 2005b. Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens. (EPA/630/R-03/003F). Washington, D.C.

United States Environmental Protection Agency (USEPA). 2016. Evaluation of the inhalation carcinogenicity of ethylene oxide (CASRN 75-21-8): In support of summary information on the Integrated Risk Information System (IRIS). EPA/635/R-16/350Fa. Washington, DC, U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment.

Valdes-Flores, C, Sielken Jr, R, Teta, J. 2010 Quantitative cancer risk assessment based on NIOSH and UCC epidemiological data. Regulatory Toxicology and Pharmacology. 56:312-320.

Valdez-Flores, C, Sielken Jr, R, Teta, J. 2011 Quantitative cancer risk assessment for ethylene oxide inhalation. Arch Toxicol. 85:1189-1193.

Valdez-Flores, C, Sielken Jr, R. 2013. Misinterpretation of categorical rate ratios and inappropriate exposure-response model. Regulatory Toxicology and Pharmacology. 67:206-214.

van Sittert, NJ, Beulink, GD, van Vliet, E, van der Waal, H, 1993. Monitoring occupational exposure to ethylene oxide by the determination of hemoglobin adducts. Environ Health Perspect 99:217-220.

World Health Organization (WHO). 2003. Concise international chemical assessment document: Ethylene oxide. In Concise International Chemical Assessment: Ethylene Oxide. Geneva, Switzerland: World Health Organization, International Programme on Chemical Safety.

Wong, O, Trent, L. 1993. An epidemiological study of workers potentially exposed to ethylene oxide. British Journal of Industrial Medicine. 50:308-316.

Appendix 1 Systematic Review and Evidence Integration

A1.1 Problem Formulation and Protocol

Problem formulation identifies and defines the causal questions and describes the extent of the evaluation. These questions structured the systematic review for EtO:

- What are the physical and chemical properties of EtO?
- What is the critical effect following exposure to EtO?
- Are there sensitive subpopulations?
- What is the mode of action (MOA)?
- Does route of exposure play a role?
- Is EtO carcinogenic, and if so, is it carcinogenic by a specific route of exposure?

Protocol development is another important aspect in the initial process. A protocol is typically developed around a PECO statement: Populations, Exposure, Comparator/Control, and Outcomes. These identifiers are used to lay out the framework for the literature search and inclusion/exclusion criteria. The PECO statement for EtO followed these criteria:

Table 17: PECO Statement used by the TCEQ to Develop Toxicity Factors for EtO

Population	General human population and any relevant sensitive subpopulations, animals, and vegetation
Exposure	Exposure to EtO, surrogates with demonstrated similar MOAs, and any identified metabolites
Comparator/ Control	Populations exposed to concentrations below the concentration that causes the most sensitive critical effect
Outcome(s)	The most sensitive critical effect directly related to EtO exposure

The protocol used for the systematic review and the development of toxicity factors for EtO is as follows:

1. Identify the chemical of interest and define the causal questions
2. Conduct a systematic review
    a. Conduct a systematic literature search
    b. Identify the inclusion/exclusion criteria
    c. Extract the relevant data from each data stream (human, animal, mechanistic)
    d. Assess the study quality and conduct a risk of bias analysis
    e. Weigh the evidence in each data stream and then integrate the evidence across the data streams

f. Rate the confidence in the evidence

3. Derive toxicity factors (TCEQ 2015)
    a. Review the essential data, including chemical/physical properties and selected key studies from the systematic review
    b. Conduct MOA analysis
    c. Choose the appropriate dose metric considering toxicokinetics and MOA
    d. Select critical effect, based on human equivalent exposure considering each key study
    e. Extrapolate from the adjusted POD to lower exposures based on MOA analysis

A1.2 Systematic Literature Review and Study Selection

As a first step, publically available databases were searched using explicitly stated search criteria. Please see TCEQ (2015) for a list of available databases that were searched. The search terms used in literature review for EtO, along with the number of results from PubMed, are found in Table 18. Additional references were also identified using the reference sections from some of the selected studies. This literature review was conducted in December 2018, and therefore studies published after this date were not available at the time of the review.

Table 18: Search Strings used in the Literature Review of EtO

Search Term/String	PubMed Results
ethylene oxide	9,626
"ethylene oxide"	7,478
"ethylene oxide" OR oxirane	10,374
"ethylene oxide" OR oxirane OR 75-21-8	10,374

These 10,374 studies were imported into the desktop application SWIFT-Review by Sciome and briefly searched to ensure that the key studies used in several other reviews were present in the data set. The data set was further narrowed down using the tag levels created by the SWIFT-Review software. The tags used and the number of studies that each tag removed can be found in Table 19.

Table 19: SWIFT-Review Tags and Results

Data Set/Tag	Number of Studies
Initial PubMed Search	10,374
Tag – Health Outcomes, any (excluded studies with no tag)	7,468
Tag – Evidence Stream, any (excluded studies with no tag)	4,914
Tag – MeSH Chemicals, only Ethylene Oxide (excluded everything else)	1,520

Additionally, several governmental and private sector organizations were searched for published literature and toxicity values for EtO (Table 20), and the available documents along with their relevant references were added to the pool of selected material as needed.

Table 20: Available Reviews and Inhalation Toxicity Values for EtO

Organization	Year	Toxicity Value
Agency for Toxic Substances and Disease Registry (ATSDR) Toxicological Profiles	1990	Intermediate MRL*
Integrated Risk Information System (IRIS) USEPA	2016	Inhalation Unit Risk
Office of Environmental Health Hazard Assessment (OEHHA) CalEPA	2000	Chronic REL* Inhalation Slope Factor

MRL – minimal risk level, REL – reference exposure level

Following this initial review, specific inclusion and exclusion criteria were used to narrow down the pool of available data. The criteria along with examples of the kinds of studies that were excluded can be found in Table 21.

Table 21: Inclusion/Exclusion Criteria used in the Review of EtO

Study Type	Inclusion Criteria	Exclusion Criteria
General	Complete study available for review	- Only abstract is available - Study in a language other than English - Unpublished report/unable to retrieve
	Study contains original data or utilizes existing data in a novel way	- Study is a review article or meta-analysis - Study comments on a previous method without providing a sufficient alternative
	Exposure concentration is known or can be reasonably estimated	- Exposure concentration unknown - Exposure environment/conditions unsuitable to concentration estimation
	Study examines effects related to chemical exposure	- Study measures concentration in air, factories, etc. - Study does not examine health effects
	Study focused on the chemical of concern	- Study examined mixture effects - Study on treatment following EtO exposure
	Route of exposure is relevant to exposure and toxicity factor development	- Exposure through i.v., i.p., or subcutaneous injection - Study examining oral or dermal exposure
Animal	Relevant animal model and endpoints examined	- Study used non-mammalian animal models - Endpoint studied not relevant to human health - Endpoint not applicable to toxicity factor development
	Appropriate study populations and methods were used	- Study lacked appropriate numbers or doses - Exposure method unsuitable for dose-response
Human/Epi	Relevant endpoints examined	- Study focused solely on cytotoxicity - Study only measured sister chromatid exchanges (SECs), protein adducts, or chromosomal changes
	Study populations allowed for significant findings and follow ups	- Case studies examining single high-dose exposures - Studies without appropriate follow-up studies - Historical studies that have been updated

i.v. – intravenous, i.p. – intraperitoneal

Studies were then divided into groups by evidence stream (i.e. human, animal) and effect group (i.e., acute, chronic non-carcinogenic, carcinogenic). For the purposes of this DSD, only the human carcinogenic/epidemiologic data were considered for several reasons:

1. In order to expedite the process, it was decided that only a health-based chronic carcinogenic toxicity factor would be derived for EtO in this DSD. Other toxicity factors (i.e. health- and welfare-based acute and chronic non-carcinogenic) may be evaluated at a later date with an additional systematic review picking up where this systematic review left off.
2. Sufficient human data exist for EtO such that animal data, although used to strengthen the carcinogenicity class, would not be used to derive a chronic carcinogenic toxicity factor. TCEQ (2015) states that in general, human data are preferred over animal data when developing toxicity factors.
3. Similarly, mechanistic data remain supportive (e.g., MOA), but not useful as a basis in the derivation of a chronic carcinogenic toxicity factor.
4. And finally, human data looking solely at cytotoxicity, sister chromatid exchanges, or chromosomal abnormalities were considered useful in developing the MOA of EtO, but not useful as a basis for derivation of a health-based toxicity factor.

After full text review and screening with the inclusion/exclusion criteria listed above, eight human carcinogenic studies were identified for further use in this systematic review. Several human studies (directly or indirectly related to carcinogenicity) were reviewed and later excluded due to various reasons (Table 22).

Table 22: Excluded Human Studies Related to Carcinogenicity

Reason for Exclusion	Study	
No exposure or dose-response information available to directly derive a toxicity factor (Not useful in the development of a carcinogenic-based toxicity factor)	Ambroise et al., 2005 Austin and Sielken, 1988 Bisanti et al., 1993 Coggon et al., 2004 Fondelli et al., 2007 Gardner et al., 1989 Greenburg et al., 1990 Greife et al., 1988 Hagmar et al., 1991 Kardos et al., 2003	Kiesselbach et al., 1990 Kiran et al., 2010 Kirman and Hays, 2017 Morgan et al., 1981 Mosavi-Jarrahi et al., 2009 Norman et al., 1995 Olsen et al., 1997 Swaen et al., 1996 Wong and Trent, 1993
Follow up study available	Greenberg et al., 1990 Hagmar et al., 1995 Hogstedt et al., 1979a Hogstedt et al., 1986	Stayner et al., 1993 Steenland et al., 1991 Teta et al., 1993
Review, methods, or case study	Hogstedt et al., 1979b Hornung et al., 1994 Kita, 1991 Shore et al., 1993 Sielken and Valdez-Flores, 2009a	Sielken and Valdez-Flores, 2009b Steenland et al., 2011 Valdez-Flores et al., 2011 Valdez-Flores and Sielken, 2013

A1.3 Data Extraction

Each of the identified studies was reviewed in detail and the primary data were extracted for potential use in the development of the chronic carcinogenic toxicity factor in this DSD (Table 23).

Table 23: Data Extraction from Epidemiological Studies

Study (cohort)	Size	Exposure Measurement	Tumor Type(s)	Notable Results[1]	Notes
Hogstedt 1988 (Swedish, chemical)	539 m 170 f	Years of employment, 1-9 years, ≥ 10 years	Stomach	SMRs – 597, 608	Exposure estimates conducted in original study but not presented here.
			Blood/Lymphatic	SMRs – 380, 330	
			Leukemia	SMRs – 322, 880	
Kirman 2004 (NIOSH + UCC)	18,254 (NIOSH) (55% m, 45% f) 1,896 m (UCC)	ppm-years, 7.4, 64.8, 187.4, 477.7	Leukemia	POD-ED_{001} estimated at 265 ppm-years, URFs: linear $4.5x10^{-7}$ /µg/m^3 Quadratic $4.5x10^{-8}$ /µg/m^3 (no lag or latency periods)	Concentration at $1x10^{-5}$ cancer risk: Linear – 22 µg/m^3 (12 ppb) Quadratic – 222 µg/m^3 (120 ppb) Nonlinear – 37 µg/m^3 (21 ppb)
Mikoczy 2011 (Swedish, sterilant)	862 m 1,309 f	ppm-years, 0-0.13, 0.14-0.21, ≥ 0.22	Breast	SIRs – 0.52, 1.06, 1.12	Compared with/out 15-year latency and between follow-ups, healthy worker effect likely influenced results
			LHN	SIRs – 1.35, 1.32, 1.08	
Steenland 2003 (NIOSH)	7,576 f (5,139 f interviewed)	ppm-days, 0, >0-647, 647-2026, 2026-4919, 4919-14620, 14620+	Breast (Compared to US population)	SIRs – 0.88, 0.77, 0.77, 0.94, 0.83, 1.27 (15-year lag, cumulative)	Subset of the NIOSH cohort, multiple other comparisons presented, including cumulative, categorical, and log cumulative exposure, positive trends for continuous exposure, duration of exposure, and log of cumulative exposure. Overall SMR for NIOSH cohort for breast cancer is 0.99. Exposure-response analysis showed highest group SMR of 1.27, with 20-year lag increased to 2.07 (95% CI: 1.0-3.54)
			Breast (Compared to study population, whole cohort)	Odds Ratios – 1.00, 1.07, 1.00, 1.24, 1.17, 1.74* (15-year lag, categorical, cumulative)	
			Breast (Compared to study population, only interviewed cohort)	Odds Ratios – 1.00, 1.06, 0.99, 1.24, 1.42, 1.87* (15-year lag, categorical, cumulative)	

Study (cohort)	Size	Exposure Measurement	Tumor Type(s)	Notable Results[1]	Notes
Steenland 2004 (NIOSH)	7,645 m 9,885 f	ppm-days, 0, >0-1199, 1200-3679, 3680-13499, 13500+	NHL	SMRs – 2.09, 0.61, 0.88, 0.79, 2.37* m, 10-year lag, cumulative	Multiple other comparisons presented, including cumulative, categorical, and log cumulative exposure, 10, 15, and 20-year lag, positive trend for lymphoid tumors
		ppm-days, 0, >0-646, 647-2779, 2780-12321, 12322+	Breast	SMRs –0.80, 1.05, 1.01, 1.15, 2.07* f, 20-year lag, cumulative	
Swaen 2009 (UCC)	2,063 m	ppm-years, 0-15, 15-65, 65+	None	Authors state no long-term carcinogenic effects associated with EtO exposure	Healthy worker effect likely influenced results, cohort experienced more than twice the average estimated cumulative exposure compared to NIOSH cohort
Teta 1999 (multiple reviewed, dose-response done for NIOSH and UCC)	Multiple, meta-analysis 8,214 m & 10,040 f (NIOSH) 1,896 m (UCC)	ppm-years, 0, 0-33, 33-125, 125-285, >285	Lymphoid (lymphocytic leukemia and NHL)	Added Risk (environmental) UCC – none NIOSH – 10^{-8} – 10^{-5} /ppb	Compared 0 and 10-year latency, and 0 and 5y lag periods, POD-ED_{001} values ranged from 0.81-1.58 ppm assuming a 10-year latency and a 5-year lag period. POD-ED_{001} of 0.81 ppm gives a URF of 0.12/ppm, and a concentration at $1x10^{-5}$ cancer risk of 0.083 ppb (0.15 µg/m^3)
			Leukemia	Added Risk (environmental) UCC – 10^{-12} – 10^{-6} /ppb NIOSH – 10^{-15} – 10^{-6} /ppb	
Valdez-Flores 2010 (NIOSH + UCC)	7,634 m & 9,859 f (NIOSH) 2,063 m (UCC)	ppm-days, dose ranges varied by endpoint	Examined 12 cancer endpoints in 6 subcohorts	No statistically significant increases in SMRs, trends, cumulative continuous, or categorical exposure.	No heterogeneity between dose-response models of the two major cohorts and the pooled study, combining increases the power.

[1] Due to space constraints, only notable results are presented here. See individual studies for a more in-depth review.

* Denotes significance, confidence interval does not include 1

SMR – Standardized mortality ratio, SIR – Standardized Incidence Ratio, NHL – Non-Hodgkin's Lymphoma, LHN – Lymphohematopoietic Neoplasms, m – males, f – females

A1.4 Study Quality and Risk of Bias (ROB)

Each of the selected studies was evaluated for study quality and ROB based on a number of attributes determined prior to this review. For this review, study quality methods were adapted from the USEPA version of the Health Assessment Workspace Collaboration (HAWC) online software. For epidemiology studies, seven evaluation domains are used to critically assess different aspects of study design and conduct relating to reporting, risk of bias and study sensitivity. Each domain receives a score of Good, Adequate, Deficient, Critically Deficient, or Not Reported, and once all domains are evaluated, a confidence rating of High, Medium, or Low confidence or Uninformative is assigned to each study. The evaluated domains and explanations can be found in Table 24, while the general guidance for scoring each of the studies can be found in Tables 25 and 26.

Table 24: Study Quality Domains for Epidemiology Studies (taken from HAWC)

Domain	Study Design Questions and Aspects
Selection and Performance/ Participant Selection	**Is there evidence that selection into or out of the study (or analysis sample) was jointly related to exposure and to outcome?** Study design, where and when was the study conducted, and who was included? Recruitment process, exclusion and inclusion criteria, type of controls, total eligible, comparison between participants and nonparticipants (or followed and not followed), final analysis group. Does the study include potential vulnerable/susceptible groups or life stages?
Exposure Methods/ Measures	**Does the exposure measure reliably distinguish between levels of exposure in a time window considered most relevant for a causal effect with respect to the development of the outcome?** Source(s) of exposure (consumer products, occupational, an industrial accident) and source(s) of exposure data, blinding to outcome, level of detail for job history data, when measurements were taken, type of biomarker(s), assay information, reliability data from repeat measures studies, validation studies.
Outcome Methods/Results Presentation	**Does the outcome measure reliably distinguish the presence or absence (or degree of severity) of the outcome?** Source of outcome (effect) measure, blinding to exposure status or level, how measured/classified, incident versus prevalent disease, evidence from validation studies, prevalence (or distribution summary statistics for continuous measures).
Confounding	**Is confounding of the effect of the exposure unlikely?** Background research on key confounders for specific populations or settings; participant characteristic data, by group; strategy/approach for consideration of potential confounding; strength of associations between exposure and potential confounders and between potential confounders and outcome; degree of exposure to the confounder in the population.

Domain	Study Design Questions and Aspects
Analysis	**Does the analysis strategy and presentation convey the necessary familiarity with the data and assumptions?** Extent (and if applicable, treatment) of missing data for exposure, outcome, and confounders, approach to modeling, classification of exposure and outcome variables (continuous versus categorical), testing of assumptions, sample size for specific analyses, relevant sensitivity analyses.
Selective Reporting	**Is there concern for selective reporting?** Are results presented with adequate detail for all the endpoints of interest? Are results presented for the full sample as well as for specified subgroups? Were stratified analyses (effect modification) motivated by a specific hypothesis?
Sensitivity	**Are there concerns for study sensitivity?** What exposure range is spanned in this study? What are the ages of participants (e.g., not too young in studies of pubertal development)? What is the length of follow-up (for outcomes with long latency periods)? Choice of referent group and the level of exposure contrast between groups (i.e., the extent to which the 'unexposed group' is truly unexposed, and the prevalence of exposure in the group designated as 'exposed'). Is the study relevant to the exposure and outcome of interest?
Overall Study Confidence	**Once the evaluation domains have been classified, these ratings will be combined to reach an overall study confidence classification of High, Medium, Low, or Uninformative.** This classification will be based on the classifications in the evaluation domains and will include consideration of the likely impact of the noted deficiencies in bias and sensitivity on the results.

Table 25: Study Quality Domain Scoring

Score	Reasoning
+ +	**Good** – Study meets or exceeds domain properties, may have minor deficiencies but none that would affect the outcome of the study or the development of toxicity factors.
+	**Adequate** – Study meets most of the domain properties, may have some deficiencies but none are severe nor are expected to have a serious effect on the development of toxicity factors.
-	**Deficient** – Study has one or more deficiencies that are likely to affect the outcome of the study or the development of toxicity factors, but development may still occur with some added uncertainty.
- -	**Critically Deficient** – Study has serious deficiencies that would severely inhibit the development of toxicity factors. These studies are typically classified as "uninformative" unless a detailed explanation otherwise is provided.
NR	**Not Reported** – Domain properties are not provided in the study or referred to in previous author's studies. Depending on the domain and type of study, these studies should be carefully considered prior to use.

Table 26: Study Quality Confidence Rating Scoring

Score	Reasoning
+ +	**High** – Overall a well conducted study, no serious deficiencies identified, no concern for issues with sensitivity or ROB, most domains should be scored good or adequate.
+	**Medium** – Some deficiencies may be noted, but nothing that would cause significant concern for issues with sensitivity or ROB, most domains should be scored adequate.
-	**Low** – Deficiencies noted, some severe, and some concern over bias or sensitivity that may impact the assessment, study has domains that scored deficient.
- -	**Uninformative** – Severe deficiencies that would seriously impact the assessment, study is typically unusable for toxicity factor development without a detailed explanation. Any study with a domain listed as "Critically Deficient" should be considered for this category.

Scoring for each of the included studies can be found in Table 27. Each reviewer scored the included studies independently, then came together to agree on a single score for each domain/study (individual scoring not shown).

Table 27: Study Quality and ROB Scoring Visual

Domain/Study	Hogstedt 1988	Kirman 2004	Mikoczy 2011	Steenland 2003	Steenland 2004	Swaen 2009	Teta 1999	Valdez-Flores 2010
Selection and Performance/Participant Selection	+	+ +	+	+	+ +	+	+ +	+ +
Exposure Methods/Measures	-	+	-	+	+	-	+	+
Outcome Methods/Results Presentation	+	+	+ +	+	+	+	+	+ +
Confounding	-	+	-	+ +	+	+	+	+
Analysis	+	+	+	+	+ +	+	+	+ +
Selective Reporting	+	+	+	+	+	+	+	+
Sensitivity	-	+	-	+	+	+	+	+
Overall Study Confidence	-	+	+	+	+	+	+	+

A1.5 Evidence Integration

After addressing the study quality and ROB for each of the selected studies, the primary information from each of the studies was compiled together and each study was assessed for use as a key, supporting, or informative study (Table 28).

Table 28: Evidence Integration Table for Human Studies

Study	Cohort	Type	Reasoning
Hogstedt 1988	Swedish chemical workers	Informative	- Relatively small cohort with little information on co-exposures - Exposure concentrations or estimations not provided - Primary cohort to show increased leukemia mortality rates - Also presented increased stomach and blood/lymphatic cancer
Kirman 2004	NIOSH + UCC	Supporting	- Combined data from two largest cohorts and examined leukemia and lymphoid tumor mortality data - Provided results for several different extrapolation methods - Selected a single outcome and POD to carry through
Mikoczy 2011	Swedish sterilant workers	Informative	- Relatively small cohort with little exposure information presented - Healthy worker effect likely influenced the results - Non-significant increases in leukemia, NHL, and lymphohematopoietic cancer mortality - Significant increases in the rate ratios of breast cancer in the two highest exposure groups
Steenland 2003	NIOSH (females only)	Informative	- Subset of the largest cohort study available, additional nested case-control using subjects who answered personal interviews - Examined breast cancer mortality and incidence data - Positive trend for increased incidence, but not significantly increased
Steenland 2004	NIOSH	Supporting	- Update to the largest EtO-exposed cohort data available - Focused mainly on hematopoietic and breast cancers, and examined various exposure variables and lag periods - No significantly increased cancer incidences, but a positive trend observed for lymphoid tumors (males, 15-year lag)

Study	Cohort	Type	Reasoning
Swaen 2009	UCC	Supporting	- Although a relatively smaller cohort, the strength of the update was made up for in the length of follow-up and number of deaths - Little to no exposure monitoring data available, estimates made from work history - Examined a wide array of cancer types but no lag/latency periods - No cancer associations observed, likely influenced by the healthy worker effect
Teta 1999	Meta-analysis, NIOSH, UCC	Supporting	- Very basic meta-analysis of 10 EtO cohorts but lacked dose-response data, detailed analysis on individual NIOSH and UCC cohorts only - Examined lymphoid and leukemia rates with various lags and latency periods and control groups using Poisson regression - UCC cohort showed no added risk, while NIOSH cohort predictions were in the range of 10^{-7} to 10^{-5} at 1 ppb environmental exposures
Valdez-Flores 2010	NIOSH + UCC	Key	- Combined most recent data from the UCC and NIOSH cohorts - Examined 12 cancer endpoints (breast, leukemia, lymphoid, etc.) and 6 sub-cohorts (NIOSH males, females, UCC males, etc.) using Cox proportional analyses without latency/lag periods - No statistically significantly increasing SMRs or trends in any of the cancer endpoints examined

After final review of the included studies, the Valdez-Flores et al. (2010) study had the most thorough and complete analysis (e.g., data from both the NIOSH and UCC cohorts, multiple cancer endpoints examined) and was therefore selected as the key study. While the Valdez-Flores et al. (2010) study also utilized a default lifetime duration (70 years) consistent with TCEQ guidance (TCEQ 2015), there were aspects that were not ideal, such as the lack of exposure lags. So rather than select a POD from the key study, the Toxicology, Risk Assessment, and Research Division (TRARD) selected data from both cohorts evaluated in the study (i.e., the NIOSH and UCC cohorts) as the key epidemiological data and conducted an independent assessment using the same approach but with supplemental analyses (e.g., the evaluation of various exposure lags). Selection of data from the NIOSH and UCC cohorts as the key epidemiological data and use of specific, TCEQ-directed dose-response assessment analyses

(rather than selection of a study POD) provide the best basis for a carcinogenic assessment of EtO for several reasons:

1. Both the NIOSH and UCC cohorts have adequate size, exposure information, and follow-up, making consideration of all the data ideal for toxicity factor development (e.g., weight of evidence, more analyses to consider).
2. The Valdez-Flores et al. (2010) study makes use of the Cox Proportional Hazard model, a standard model that the TRARD has used previously in dose-response assessments (also considered by USEPA 2016).
3. Although Valdez-Flores et al. (2010) did not include exposure lag results in their publication, supplemental analyses involving a reassessment of the data using various exposure lags allow for the consideration of even more assessment results in the DSD.
4. Additionally, since published in 2010, an update to the UCC data through 2013 has become available to the first author of the Valdez-Flores et al. (2010) study (submitted for publication), who the TCEQ contracted with to perform supplemental analyses; consequently, results from the new study update with a longer follow-up period can also be included in the DSD.
5. Unlike USEPA (2016) that uses a lifetime exposure duration value of 85 years, the TCEQ-directed dose-response analyses use a standard default of 70 years consistent with TCEQ guidance (TCEQ 2015).
6. And finally, conducting these new analyses will allow for the appropriate consideration of model fit to the individual data (rather than the categorical data) for the model assessment ultimately selected by the TRARD.

A1.6 Confidence Rating

Table 29 provides scoring criteria to rate the confidence and uncertainty for each aspect or element of the toxicity assessment. The table provides the name of the element and the magnitude of the confidence in each element using a qualitative ranking system of low, medium, or high confidence. Table 30 displays the overall confidence in the EtO carcinogenic assessment. Once the noncarcinogenic assessments are completed for EtO, the confidence rating will be updated to cover the entire assessment.

Table 29: Confidence Scoring Criteria for EtO Carcinogenic Assessment

Element	Low	Medium	High
Database Completeness	Only a single study or a few low-quality studies were available.	Several studies were available, but some important studies were missing.	Several high-quality studies were available to select from.
Systematic Review	A systematic approach was not used.	A systematic approach was considered and some methods were applied, but a full review was not conducted	A systematic approach was used in study evaluation and clear criteria were established for judgment
Key Study Quality	Selected study has deficiencies, but was still considered useful	Selected study was reasonably well done but some restrictions must be considered	Selected study was well done and can be used without restriction
Critical effect	Critical effect or dose-response curve was moderate to severe. MOA information was not available.	Critical effect was moderate; other studies were deemed necessary to determine the critical effect.	Critical effect was minimal, or the confidence in the critical effect was high. MOA information was available.
Relevance of Critical Effect	Critical effect was only presumed to be relevant for the general population; MOA was not known for the critical effect.	Critical effect appeared to be relevant for the general population. MOA was known for the critical effect and possibly relevant to humans.	Critical effect based on a human study or matches observed human experience; MOA was well understood so critical effect was assumed relevant.
Point of Departure (POD)	Many uncertainties exist in POD; only a few dose groups; no dose-response modeling was used.	Some uncertainty exists in POD; few dose groups; difference between confidence limits was large.	Basis for POD well understood; multiple dose groups, dose-response modeling was conducted.
Sensitive Populations	Many uncertainties on sensitive population(s) existed and were not addressed.	Information on sensitive population(s) was not known but default procedures are presumed to be conservative.	Human data on sensitive populations were available and uncertainties were addressed.
Peer Review	Limited or no peer review; disregarded comments would significantly change risk value; no independent check	Adequate peer review. Most substantive comments addressed; disregarded comments would not significantly change value	High quality panel peer review with appropriate experts; all substantive comments addressed as per independent check
Toxicity Value Comparison	Relevant risk values show a greater than 10-fold difference without justification.	Some relevant risk values agreed within 3-fold of each other, others disagreed within 10-fold without justification.	All relevant risk values agreed within 3-fold of each other or there was sufficient justification for differences.

Table 30: Confidence in the Toxicity Assessment

Element	Score	Basis
Database Completeness	Medium	- Several occupational cohorts (i.e., preferred human data) and animal studies available - Evidence of carcinogenic effects found in both human epidemiological and animal studies - However, estimated exposures are based on incomplete information, are remarkably high, and are not in/near lower range of interest (i.e., not environmentally relevant)
Systematic Review	High	- Systematic review conducted
Key Study Quality	High	- Well-conducted study of two cohorts and multiple cancer endpoints with standard Cox proportional hazards modeling but lacked the use of a lag period - Reassessment of these key epidemiological data utilizing multiple exposure lags and new UCC cohort data allowed for informative supplemental and updated analyses
Critical effect	Low	- Human data not conclusive despite remarkably high exposure (e.g., results vary between studies) - Model (slope > 0) not statistically significantly different than the null model (slope = 0) at the 5% significance level
Relevance of Critical Effect	Medium	- Assumed relevant although general population exposed to levels orders of magnitude lower than the occupational study wherein lymphoid cancer was statistically increased only in the highest cumulative exposure group
Point of Departure (POD)	High	- Cox Proportional Hazard model used - Modeling results demonstrated to be predictive
Sensitive Populations	Medium	- No specific data on sensitive subpopulations - Default ADAFs were applied to account for potentially increased susceptibility in children due to early-life exposure
Peer Review	Medium	- DSD proposed for public comment and reviewed by a consulting academic statistician and subject matter expert in regard to potential statistical issues at TCEQ's direction
Toxicity Value Comparison	High	- TCEQ Chronic ESL based on lymphoid cancer mortality is 4,000 times higher than the USEPA value based on lymphoid/breast cancer incidence at the same excess risk level (1E-05)

Element	Score	Basis
		- TCEQ's approach is supported by multiple lines of evidence as discussed in the DSD, whereas USEPA's non-standard approach is not - Extensive comparisons, calculations, and explanations as to the differences and errors in USEPA's methods are included in the DSD (e.g., USEPA's model assessment is demonstrated to be statistically significantly over-predictive)

Confidence Scoring Summary			
Not Evaluated	**Low Confidence**	**Medium Confidence**	**High Confidence**
	Critical Effect	Database Completeness	Systematic Review
		Relevance of Critical Effect	Key Study Quality
		Sensitive Populations	Point of Departure
		Peer Review	Toxicity Value Comparison

Appendix 2 Additional Analysis of Kirman and Hays (2017) Data

Ciriaco Valdez-Flores, Ph.D., P.E.
Professor of Practice
Industrial and Systems Engineering
Texas A&M University
College Station, TX 77843

A statistically significant change in the true HEV mean based on the meta-analysis by Kirman and Hays (2017) for the nonsmoker/passive smoke-exposed ("unexposed") population (n=661) was determined at the 0.05 significance level (α). The mean (μ_L) HEV from the study was 21.1 pmol/g Hb with a standard deviation (σ_L) of 14.6 pmol/g Hb. This mean and standard deviation correspond to a lognormal distribution. That is, the natural logarithm of HEV (ln(HEV)) follows a normal distribution with mean (μ_N) and standard deviation (σ_N). The values corresponding to μ_N and σ_N are determine from μ_L and σ_L using the following relationship:

$$\mu_N = \ln\left(\frac{\mu_L^2}{\sqrt{\mu_L^2 + \sigma_L^2}}\right)$$

and

$$\sigma_N = \sqrt{\ln\left(\frac{\mu_L^2 + \sigma_L^2}{\mu_L^2}\right)}$$

Once the parameters of the distribution of the ln(HEV) are calculated, then the null hypothesis to test is:

$$H_0: \mu_N^* = \mu_N$$

where μ_N^* is the true mean of the ln(HEV) of the distribution of the sample.

The significance level (α=0.05) and the sample size (n=661) have to be specified to determine the maximum (μ_N^{*+}) and minimum (μ_N^{*-}) values of the sample average of ln(HEV) that result in statistically significant changes from the mean of the ln(HEV) (μ_N). These maximum and minimum values are calculated as follows:

$$\mu_N^{*+} = \mu_N + Z_{(1-\alpha)} \frac{\sigma_N}{\sqrt{n}} \text{ and } \mu_N^{*-} = \mu_N - Z_{(1-\alpha)} \frac{\sigma_N}{\sqrt{n}}$$

Note that the significance level (α) is a one-sided significance level. That is, if it is only interesting to detect an increase in the ln(HEV) then only the maximum is relevant.

Given μ_N^{*+} and μ_N^{*-}, the corresponding maximum and minimum sample geometric means that result in statistically significant changes from the mean ln(HEV) are $\mu_G^{*+} = \exp\{\mu_N^{*+}\}$ and $\mu_G^{*-} = \exp\{\mu_N^{*-}\}$, respectively.

The maximum (μ_L^{*+}) and minimum (μ_L^{*-}) mean of the log-normal distribution corresponding to the HEV samples can be calculated as follows:

$$\mu_L^{*+} = \exp\left\{\mu_N^{*+} + \frac{\sigma_N^2}{2}\right\} \quad \text{and} \quad \mu_L^{*-} = \exp\left\{\mu_N^{*-} + \frac{\sigma_N^2}{2}\right\}$$

Thus, the equivalent multiple of the mean that results in a statistically significant increase or decrease in HEV are given by μ_L^{*+}/μ_L and μ_L^{*-}/μ_Lrespectively. Similarly, the increase or decrease in the average HEV that result in a statistically significant change in the mean HEV are given by $\mu_L^{*+} - \mu_L$ and $\mu_L^{*-} - \mu_L$, respectively.

Accordingly, the equation $\mu_L^{*+} - \mu_L$ ultimately provides the statistically significant increase in the mean HEV for the population (n=661) at the specified significance level (α=0.05).

The endogenous equivalent concentration in air that results in a statistically significant increase or decrease in the mean of ln(HEV) then is equal to $\mu_L^{*+}/10.9$ or $\mu_L^{*-}/10.9$, respectively. The increase or decrease in the endogenous equivalent concentration that results in an increase or decrease in the mean of the LN(HEV) are then equal to $(\mu_L^{*+} - \mu_L)/10.9$ or $(\mu_L - \mu_L^{*-})/10.9$, respectively.

Accordingly, the equation $(\mu_L^{*+} - \mu_L)/10.9$ ultimately provides the continuous air exposure concentration corresponding to the calculated statistically significant increase in mean HEV for the population (n=661) at the specified significance level (α=0.05).

Per the above approach, at a significance level (α) of 0.05 for the nonsmoker/passive smoke-exposed population (n=661) in the Kirman and Hays (2017) meta-analysis, an increase of 0.861 pmol/g in the HEV mean (from 21.1 to 21.961 pmol/g Hb) would be statistically significant (p=0.05). A continuous ethylene oxide air exposure concentration ≥0.079 ppb is calculated to be sufficient to induce this increase in the HEV.

Appendix 3 Reality Check of Epidemiological Exposure-Response Model Results for EtO and Lymphoid Cancer Mortality

USEPA fit several alternative parametric models for lymphoid cancer mortality in the NIOSH cohort and compared the predicted rate ratios by each model with non-parametric estimates of rate ratios. USEPA used the visual comparison of the parametric and non-parametric rate ratios as one of their criteria to select their parametric model. A more robust comparison is to see how reasonable the parametric models are when comparing what the models predict in terms of lymphoid cancer deaths versus the actual number of deaths in the NIOSH cohort. A good parametric model should predict the observed number of lymphoid cancer deaths with some confidence (e.g., the observed number of lymphoid cancer deaths in the NIOSH cohort should be inside a 95% confidence interval of the estimated number of lymphoid cancer deaths).

Here, some of the USEPA models and one model developed by Sielken & Associates (S&A) were used to check whether the models were reasonable; that is, whether the models predicted within a margin of error, the number of lymphoid cancer deaths in the NIOSH cohort. The estimated number of lymphoid cancer deaths for a specific model for the rate ratios were calculated using age-, sex-, race-, and calendar-year specific background hazard rates. Sections C and D of this appendix illustrates how the calculations to determine the number of expected deaths for each model were performed with methodology used in the calculation of standard mortality ratios (SMRs). The SMR is a measure that shows the ratio of observed to expected number of deaths in the cohort. Similarly, the 100(1-α)% confidence interval on the SMR is a confidence interval on the ratio of observed to expected number of deaths in the cohort.

Herein, the inverse of the SMR is used as a measure of over-prediction or under-prediction of the actual number of observed deaths. That is, the inverse of the SMR (SMR^{-1}) is the ratio of expected to observed number of deaths. Similarly, the inverse of the confidence limits of the 100(1-α)% confidence interval on the SMR result in a 100(1-α)% confidence interval on the inverse of the SMR. In turn, using the SMR^{-1} and its 100(1-α)% confidence interval, a 100(1-α)% confidence interval on the expected or predicted number of deaths can be easily calculated. Using this confidence interval on the predicted number of deaths can then be compared with the observed number of deaths. If the observed number of deaths is inside the 100(1-α)% confidence interval, then the expected number and observed number of deaths are not statistically significantly different at the α% significance level. If the observed number of deaths is below the lower end or above the upper end of the 100(1-α)% confidence interval, then the expected number is statistically significantly different than the observed number of deaths at the α% significance level.

At issue is the predictiveness (or lack thereof) of the model assessments ultimately used by USEPA and the TCEQ. *There is no fairer evaluation of the predictiveness of a given model assessment than direct numerical comparisons of the specific model's predictions to the reality of the dose-response data.* **Upon performing this evaluation, the sections below show that only the log-linear model (standard Cox proportional hazards model) and the best estimates of the linear model predict the number of observed lymphoid deaths in the NIOSH cohort with 95% confidence.** *By contrast, the specific model assessment chosen by USEPA (i.e., the upper bound of the linear two-piece spline model with the "knot" at 1,600 ppm-days; 15-year exposure lag) statistically significantly over-estimates (statistically significant at the 5% significance level) the number of observed lymphoid cancer deaths* (even after restricting those models to assume zero increase in the rate ratio for cumulative exposures above the knot).

A3.1 Predicted Versus Observed Number of Lymphoid Cancer Deaths in the NIOSH Cohort

Table 31 and Figure 13 below shows the predicted number of lymphoid cancer deaths in the NIOSH cohort for male and female NIOSH workers using several different EtO exposure-response models. There are 53 lymphoid cancer deaths in the NIOSH cohort (brown horizontal line in Figure 13). Several exposure-response models fit to the NIOSH data were used to estimate the number of lymphoid cancer deaths that the model would predict in the NIOSH cohort, if the fitted model were true. The maximum likelihood estimates of the model as well as the upper 95% confidence limit on the model parameters were used to obtain the predicted number of deaths. In addition to calculating the expected number of deaths predicted by each model and its upper bound on the slope, a 95% confidence interval in the predicted number of deaths was derived using a confidence interval for the ratio of the predicted to the observed number of lymphoid cancer deaths in the NIOSH cohort.

The 95% confidence intervals for the number of lymphoid cancer deaths predicted by the log-linear models and its upper bounds (Cox proportional hazards model, models 1, 2, 3, and 4) include the number of lymphoid cancer deaths actually observed (53) in the NIOSH cohort. The 95% confidence interval for the number of lymphoid cancer deaths predicted by the best estimate of the linear model (model 5) also includes the number of lymphoid cancer deaths actually observed in the NIOSH cohort, but the upper bound of the linear model (model 6) statistically significantly over-predicts the observed number of lymphoid cancer deaths.

Models 7, 8, 9 and 10 are USEPA's two-piece spline models. Every two-piece spline model estimate of the lymphoid cancer deaths in the NIOSH cohort statistically significantly over-predicts the actual number of lymphoid cancer deaths in the NIOSH cohort. For comparison purposes, Models 11, 12, 13 and 14 are USEPA two-piece spline models restrained by setting the slope after the knot equal to zero (i.e., the rate ratio increases with cumulative exposure up to the knot and stays flat after the knot). Even for these restrained two-piece spline

models, for both the MLE and 95% UCL, every model estimate of the lymphoid deaths in the NIOSH cohort statistically significantly over-predicts the actual number of lymphoid deaths in the NIOSH cohort.

Table 31: USEPA's Selected Model Assessment Statistically Significantly Over-Predicts Lymphoid Cancer Mortalities

Model	Slope Parameter (per ppm-day)	Predicted if the Model were True	100% x Ratio: Predicted / Observed	95% CI on Predicted if the Model were True
Background (No Model)	n/a	50.39	95.1%	(38.5, 67.3)
1. S&A – Loglinear – 15-yr lag (MLE)[1]	2.81E-06	52.42	98.9%	(40.1, 70.0)
2. S&A – Loglinear – 15-yr lag (95% UCL)[1] - **TCEQ Adopted**	7.17E-06	58.75	110.8%	(44.9, 78.4)
3. USEPA - Loglinear - 15-yr Lag (MLE)[1] USEPA Table 4-2	4.74E-06[2]	54.52	102.9%	(41.7, 72.8)
4. USEPA - Loglinear - 15-yr Lag (95% UCL)[1] USEPA Table 4-2	1.03E-05[3]	66.41	125.3%	(50.8, 88.7)
5. USEPA - Linear - 15-yr Lag (MLE) USEPA Table D-36	1.23E-05[4]	57.58	108.6%	(44.0, 76.9)
6. USEPA - Linear - 15-yr Lag (95% UCL) USEPA Table D-36	**4.71E-05[5]**	**77.3**	**145.8%**	**(59.1, 103.2)**
USEPA Spline Model with Knot at 1,600 ppm-days				
7. USEPA – Loglinear Spline – 15-yr lag (MLE) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**4.89E-04[6]**	**88.24**	**166.5%**	**(67.5, 117.8)**
8. USEPA – Loglinear Spline – 15-yr lag (95% UCL) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**9.08E-04[7]**	**144.15**	**272.0%**	**(110.2, 192.5)**
9. USEPA – Linear Spline – 15-yr lag (MLE) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**7.58E-04[8]**	**91.69**	**173.0%**	**(70.1, 122.4)**
10. USEPA – Linear Spline – 15-yr lag (95% UCL) – USEPA Table 4-4 Knot @ 1,600 ppm-days - USEPA Selected	**1.80E-03[9]**	**141.09**	**266.2%**	**(107.9, 188.4)**
Results using above USEPA models but assuming that slope for RR is zero after the "knot"				

Model	Slope Parameter (per ppm-day)	Predicted if the Model were True	100% x Ratio: Predicted / Observed	95% CI on Predicted if the Model were True
11. USEPA – Loglinear Spline – 15-yr lag (MLE) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**4.89E-04**	**84.59**	**159.6%**	**(64.7, 112.9)**
12. USEPA – Loglinear Spline – 15-yr lag (95% UCL) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**9.08E-04**	**141.97**	**267.9%**	**(108.5, 189.5)**
13. USEPA – Linear Spline – 15-yr lag (MLE) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**7.58E-04**	**86.39**	**163.0%**	**(66.0, 115.3)**
14. USEPA – Linear Spline – 15-yr lag (95% UCL) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**1.80E-03**	**135.19**	**255.1%**	**(103.4, 180.5)**

[**Boldface** values indicate that the model over-prediction of lymphoid cancer deaths is statistically significant.]

[1]The models used by Sielken & Associates and EPA [appearing as an appendix in USEPA (2016)] are the same models; however, USEPA did not use all of the individual data – Steenland et al. and USEPA only used a subsample of the individual data.

[2]The best estimate and standard error of the slope are 4.74E-06 and 3.35E-06, respectively.

[3]The 95% upper confidence limit on the slope is 1.03E-05 (4.74E-06 + 1.645×3.35E-06).

[4]The best estimate and standard error of the slope are 1.23E-05 and 2.12E-05, respectively. The standard error (2.12E-05) of the slopes was inferred from the upper bound on the slope (4.75E-05) given in Table D-36; that is 1.23E-0-5 = (4.71E-05 – 1.23E-05)/1.645.

[5]The 95% upper confidence limit on the slope is 4.71E-05 from Table D-36.

[6]The best estimate and standard error of the slope below the knot are 4.89E-04 and 2.55E-04, respectively. The slope and corresponding standard error after the knot are -4.86E-04 and 2.56E-04, respectively, from Tables 4-4 and D-33 log-linear with knot @ 1600 ppm-days.

[7]The slope after the knot is for the 95% upper confidence limit for the model is -9.07E-04 (-4.86E-04 - 1.645×2.56E-04, which a 95% LCL on the slope above the knot). This conservatively assumes perfect negative correlation of the slope below and after the knot. **Thus, the over-prediction may be larger than what is shown in the table.** The assumption of perfect negative correlation is consistent with the covariance values obtained by EPA for two-piece linear spline model; e.g., see footnote to Table D-36 in the appendices of EPA's report.

[8]The best estimate and standard error of the slope below the knot are 7.58E-04 and 6.32E-04, respectively. The slope and corresponding standard error after the knot are -7.48E-04 and 6.31E-04, respectively, from footnote to Table D-36.

[9]The slope after the knot is for the 95% upper confidence limit for the model is -1.79E-03 (-7.48E-04 - 1.645×6.32E-04, which a 95% LCL on the slope above the knot). This conservatively assumes perfect negative correlation of the slope below and after the knot. **Thus, the over-prediction may be larger than what is shown in the table.** The assumption of perfect negative correlation is consistent with the covariance values obtained by EPA (see footnote to Table D-36 in the appendices of EPA's report where the covariance is approximately equal to the negative of the variances for the slopes above and below the knot; i.e., covariance=-3.99E-07, Var1=3.99E-07, and Var2=3.98E-07).

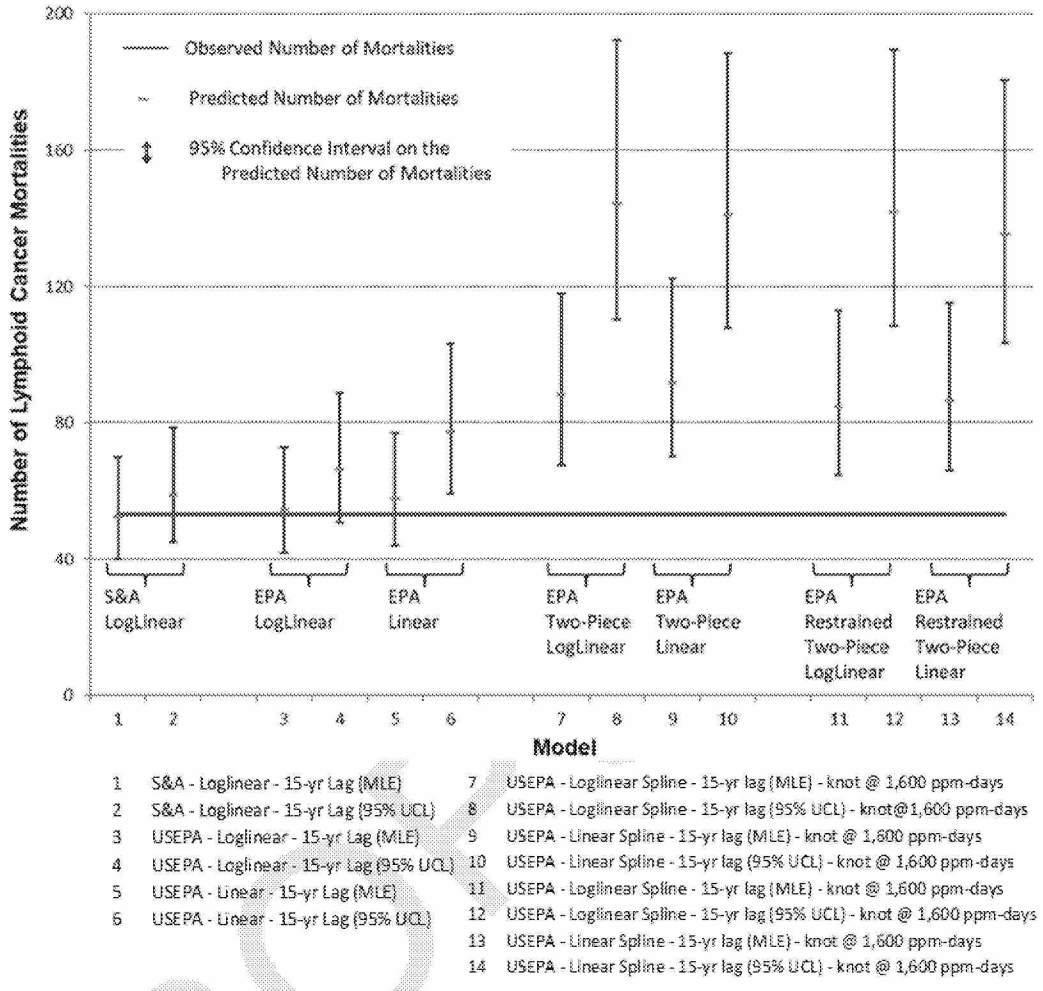

Figure 13: USEPA's Selected Model Statistically Significantly Over-Predicts Lymphoid Cancer Mortalities

A3.2 Predicted Versus Observed Number of Lymphoid Cancer Deaths in the NIOSH Cohort by Quintiles

Table 32 expands the results in Table 31 to calculate the observed and expected number of lymphoid cancer deaths in each of five quintiles. The NIOSH cohort was divided into five quintiles. A total of 53 lymphoid cancer deaths were observed in the NIOSH cohort. The first quintile included the nine NIOSH workers who died with lymphoid cancer and whose cumulative exposure to EtO (lagged 15 years) was equal to zero. Cumulative exposures to EtO lagged 15 years were defined so that quintiles 2 to 5 included the same number of lymphoid cancer deaths (11) in each quintile.

Only the best estimates of the log-linear (Cox proportional hazards) model (models 1 and 3), the linear model (model 5), and the 95% upper confidence limit of the Sielken & Associates log-linear model (model 2) predict a number of lymphoid cancer mortalities that are consistent with the number of observed deaths in each of five quintiles. USEPA's 95% UCL of the log-linear (model 4) and linear model (model 6) statistically significantly over-predict the number of the lymphoid cancer deaths in the highest exposure group.

USEPA's two-piece spline models (both the fitted models 7-10 and the restrained models 11-14) significantly over-predict the number of observed lymphoid cancer deaths at the lowest exposure quintile. *The 95% UCL of the two-piece spline models (for both the fitted models and the restrained models - models 8, 10, 12, and 14) significantly over-predict the number of observed lymphoid cancer deaths at every exposure quintile.* **More specifically, the model assessment selected by USEPA (i.e., the upper bound of the linear two-piece spline model with the "knot" at 1,600 ppm-days; 15-year exposure lag) statistically significantly over-predicts lymphoid cancer deaths for every quintile, even if the slope of the upper spline is set to zero (see Table 32 results for models 10 and 14).** The best estimates of the two-piece spline models (for both the fitted models and the restrained models - models 7, 9, 11, and 13) significantly over-predict the number of observed lymphoid cancer deaths in exposure quintiles 2 and 4 (model 9 also significantly over-predicts quintile 5).

Thus, in addition to USEPA's selected model assessment (i.e., upper bound of the linear two-piece spline model with the "knot" at 1,600 ppm-days; 15-year exposure lag) statistically significantly over-estimating the total number of observed lymphoid cancer deaths for the NIOSH cohort (141 predicted versus 53 actually observed; Table 31), their selected model also statistically significantly over-predicts lymphoid cancer deaths for every cumulative exposure group, even if the slope of the upper spline is set to zero (Table 32). By contrast, the model assessment selected by the TCEQ (i.e., upper bound of the log-linear/Cox proportional hazards model; 15-year exposure lag) is reasonably accurate, neither significantly over- or under-estimating lymphoid cancer deaths for cumulative exposure groups or for the cohort as a whole (59 predicted versus 53 observed).

Table 32: USEPA's Selected Model Statistically Significantly Over-Predicts Lymphoid Cancer Mortalities for All Cumulative Exposure Groups

Model [1]	Quintile 2*	Quintile 3	Quintile 4	Quintile 5
Observed	11	11	11	11
Background (No Model)	14.4 (8.0, 28.9)	7.9 (4.4, 15.9)	9.1 (5.1, 18.3)	7.4 (4.2, 14.9)
1. S&A – Loglinear – 15-yr lag (MLE)	14.4 (8.1, 28.9)	8.0 (4.5, 16.1)	9.4 (5.2, 18.8)	9.1 (5.1, 18.3)
2. S&A – Loglinear – 15-yr lag (95% UCL) - **TCEQ Adopted**	14.5 (8.1, 29.0)	8.1 (4.5, 16.2)	9.8 (5.5, 19.6)	15.0 (8.4, 30.0)
3. USEPA - Loglinear - 15-yr Lag (MLE) USEPA Table 4-2	14.4 (8.1, 29.0)	8.0 (4.5, 16.1)	9.5 (5.3, 19.1)	11.0 (6.2, 22.1)
4. USEPA - Loglinear - 15-yr Lag (95% UCL) USEPA Table 4-2	14.5 (8.1, 29.1)	8.2 (4.6, 16.4)	10.0 (5.6, 20.1)	**22.2 (12.4, 44.6)**
5. USEPA - Linear - 15-yr Lag (MLE) USEPA Table D-36	14.5 (8.1, 29.1)	8.2 (4.6, 16.5)	10.2 (5.7, 20.4)	13.2 (7.4, 26.5)
6. USEPA - Linear - 15-yr Lag (95% UCL) USEPA Table D-36	14.8 (8.3, 29.7)	9.0 (5.0, 18.0)	13.1 (7.3, 26.3)	**28.9 (16.2, 58.0)**
EPA Spline Model with Knot at 1,600 ppm-days				
7. USEPA – Loglinear Spline – 15-yr lag (MLE) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**19.8 (11.1, 39.7)**	17.3 (9.7, 34.7)	**20.3 (11.3, 40.7)**	19.4 (10.8, 38.9)
8. USEPA – Loglinear Spline – 15-yr lag (95% UCL) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**27.0 (15.1, 54.2)**	**33.5 (18.7, 67.3)**	**38.8 (21.7, 77.9)**	**33.3 (18.6, 66.7)**
9. USEPA – Linear Spline – 15-yr lag (MLE) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**20.9 (11.7, 42.0)**	17.6 (9.8, 35.2)	**20.8 (11.6, 41.7)**	**20.9 (11.7, 41.9)**
10. USEPA – Linear Spline – 15-yr lag (95% UCL) – USEPA Table 4-4 Knot @ 1,600 ppm-days - USEPA Selected	**29.9 (16.7, 60.0)**	**30.5 (17.1, 61.2)**	**35.8 (20.0, 71.7)**	**33.4 (18.7, 67.1)**
Results using above USEPA two-piece spline models but assuming that slope for RR is zero after the "knot"				
11. USEPA – Loglinear Spline – 15-yr lag (MLE) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**19.8 (11.1, 39.7)**	17.3 (9.6, 34.6)	**19.9 (11.1, 39.9)**	16.2 (9.0, 32.5)

Model [1]	Quintile 2*	Quintile 3	Quintile 4	Quintile 5
12. USEPA – Loglinear Spline – 15-yr lag (95% UCL) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**27.0 (15.1, 54.2)**	**33.5 (18.7, 67.2)**	**38.6 (21.6, 77.4)**	**31.3 (17.5, 62.8)**
13. USEPA – Linear Spline – 15-yr lag (MLE) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**20.9 (11.7, 42.0)**	17.5 (9.8, 35.0)	**20.1 (11.2, 40.3)**	16.4 (9.1, 32.8)
14. USEPA – Linear Spline – 15-yr lag (95% UCL) – USEPA Table 4-4 Knot @ 1,600 ppm-days	**29.9 (16.7, 60.0)**	**30.4 (17.0, 61.0)**	**35.0 (19.5, 70.2)**	**28.4 (15.9, 57.0)**

[**Boldface** values indicate that the model over-prediction of lymphoid cancer deaths for the quintile is statistically significant.]

[1] The models used to calculate the estimated number of lymphoid deaths are the same as those listed in in Table 31 and the footnotes to Table 31 apply here also. Except that the assumption of perfect negative correlation of the slopes before and after the knot in Models 8 and 10 (EPA's 95% UCL for the two-piece spline models) do not affect the predictions in quintile 2.

* Quintile 1 is the control (unexposed lagged-out) group with 9 lymphoid cancer mortalities observed and 11.5 mortalities predicted by all models with a 95% confidence interval of (6.0, 25.2), which includes the observed 9 lymphoid cancer deaths.

A3.3 Calculation of the Expected Number of Case-Specific Deaths in a Cohort Using US Background Hazard Rates

The SMR is a measure that compares the number of observed case-specific deaths in a study population with the number of case-specific deaths expected in the study population with known case-specific background death rates of a reference population. The case-specific background death rates of the reference population can adjust for calendar year, age, sex, race, and other relevant variables that may influence the case-specific death rates. The SMR is calculated using the following equation:

$$SMR = \frac{\sum_i y_{oi}}{\sum_i p_{oi} \frac{y_{ri}}{p_{ri}}}$$

where *i* is the stratum (the stratum is calendar year, age, sex, and race combination), y_{oi} is the number of observed deaths in the i-th stratum of the study group, p_{oi} is the number of observed person-years in the i-th stratum of the study group, y_{ri} is the number of deaths in the i-th stratum of the reference population, and p_{ri} is the number of person-years in the i-th stratum of the reference population.

The ratios $\frac{y_{ri}}{p_{ri}}$ are the stratum-specific mortality rates in the reference population. The SMR is then the ratio of the number of case-specific deaths in the study population $(\sum_i y_{oi})$ to the expected number of case-specific deaths in the study population $(\sum_i p_{oi}\frac{y_{ri}}{p_{ri}})$ estimated using the background case-specific death rates of the reference population. Several references have a more in-depth discussion of SMRs (e.g., Rothman 1986, Breslow and Day 1987, Checkoway, Pearce, and Crawford-Brown 1989).

The numerator in the SMR calculation is the sum of the calendar year, sex, race, and age-specific lymphoid cancer deaths in the NIOSH study $(\sum_i y_{oi})$ and is equal to the number of observed lymphoid cancer deaths. The denominator in the SMR calculation is the expected number of lymphoid cancer deaths in the NIOSH workers assuming that lymphoid was the only cause of death by using the US background lymphoid cancer mortality rates. The calendar year, sex, race, and age-specific lymphoid cancer mortality rates (y_{ri}/p_{ri}) for the US populations and the calendar year, sex, race, and age-specific person-years in the NIOSH study (p_{oi}) were used to calculate the expected number of the lymphoid cancer deaths in NIOSH workers.

An SMR greater than 1 (or 100%) implies that the number of observed deaths in the cohort is more than would be expected in a population with the same demographic characteristics as the cohort, except for potential exposures on the job. In contrast, an SMR less than 1 (or 100%) implies that the number of observed deaths in the cohort is less than would be expected in a population with the same demographic characteristics as the cohort, except for potential exposures on the job. The point estimate of the SMR cannot be used to derive statistically relevant conclusions indicating whether the observed number of deaths is greater or less than the expected number of deaths with a specific degree of confidence. Breslow and Day (1987) present the following equations that can be used to derive 100(1-α)% confidence intervals for the SMR.

$$SMR_{LCL} = \frac{Obs}{E} \times \left(1 - \frac{1}{9 \times Obs} - \frac{Z_{\alpha/2}}{3 \times \sqrt{Obs}}\right)^3$$

and

$$SMR_{UCL} = \frac{(Obs+1)}{E} \times \left(1 - \frac{1}{9 \times (Obs+1)} + \frac{Z_{\alpha/2}}{3 \times \sqrt{Obs+1}}\right)^3$$

where SMR_{LCL} is the 100(1-α/2)% lower confidence limit on the SMR, SMR_{UCL} is the 100(1-α/2)% upper confidence limit on the SMR, *Obs* is the number of observed cause-specific deaths (e.g., lymphoid cancer deaths) in the study $(i.e.,\ Obs = \sum_i y_{oi})$, *E* is the expected cause-specific deaths (e.g., lymphoid cancer deaths) derived from the reference population background rates